

TITLE

REGULATOR/PROMOTER FOR TUNABLE GENE EXPRESSION
AND METABOLITE SENSING

This application claims the benefit of United States Provisional
5 Application 60/440,965 filed January 17, 2003, which is hereby
incorporated in its entirety by reference.

FIELD OF INVENTION

The present invention relates to the fields of molecular biology and
microbiology. More specifically, this invention pertains to a novel system
10 for controlling gene expression levels that is induced by inexpensive,
environmentally friendly, small molecules.

BACKGROUND

There is a need in the field of microbial metabolic engineering for
tunable promoters and novel regulatory switches for the inducible
15 expression of heterologous proteins. The "old standard" *lac* promoter/Lac
repressor system is still widely used. However, the commonly used
inducer, isopropylthio-beta-D-galactoside ("IPTG"), is expensive and thus
not practical for large-scale bioprocesses. Additionally, higher
concentrations of IPTG increase the metabolic burden on the cell, in turn
20 reducing the maximal expression of the target gene (Donovan et al., *J. Ind.
Microbiol.* 16:145-154 (1996)). The few available alternatives also have
limitations.

One possible solution to this problem is to appropriate existing
genetic systems of transcriptional regulators to enhance heterologous
25 gene expression. A number a of transcriptional regulators are known. For
example, the LysR family of transcriptional regulators is one of the largest
groups of transcriptional regulators in prokaryotes (Schell, *Annu. Rev.
Microbiol.* 47:597-626 (1993)). Currently, there are over 80 known
members of this regulator family. Proteins having greater than 20% amino
30 acid identity with another LysR family member or having the consensus
sequence of the N-terminal region of the LysR family are considered to be
members of this regulator family. LysR family members are also
commonly found in the size range of 276 to 324 amino acids, bind to
similar DNA sequences in the absence of inducers, have promoters that
35 are located close to or overlapping those of the regulated target gene, and
most can repress their own transcriptional levels 3- to 10-fold. Activation
of the regulated target gene occurs in the presence of inducer and usually

results in a 6- to 200- fold increase in regulated target gene transcription. Regulated target genes are diverse and have numerous functions.

Recently, the gene encoded by open reading frame ("ORF") b3243 in *Escherichia coli* ("*E. coli*") has been demonstrated to function via quorum sensing (Sperandio et al., *Infect. Immun.* 70:3085-3093 (2002)). Quorum sensing is the ability of bacteria cells to communicate with one another through perception of the accumulation of signaling molecules based on bacteria cell density. The gene encoded by ORF b3243 is up-regulated via quorum sensing resulting in a 23-fold increase in transcription of the gene. The protein produced by the b3243 ORF was not purified. The gene was found to have a role in the regulation of the LEE genes involved in a type III secretion system, a pathogenecity system that serves to translocate, upon contact with eukaryotic host cells, proteins from the bacteria cytoplasm into the host cell cytoplasm. As the b3243 ORF gene is a putative regulator of the LysR family, the b3243 ORF gene itself was able to induce a four-fold induction of LEE1 transcription. No inducer of the b3243 ORF gene was identified. One in the art will appreciate that this inefficient induction prevents the b3243 ORF gene from being a viable promoter/regulator system without the identification of its inducer.

The use of promoter/reporter gene constructs is well known in the art (Serebriiskii and Golemis, *Anal. Biochem.* 285:1-15 (2000); Spergel et al., *Prog. Neurobiol.* 63:673-686 (2001); Yarranton, *Curr. Opin. Biotechnol.* 3:506-511 (1992)). Particularly, reporter systems utilizing β -galactosidase, green florescent protein, luciferase, and chloramphenicol acetyl transferase (CAT) are all commonly used in the art. Additionally, reporter systems such as *luxCDABE* are useful because this operon contains all of the genes required for bioluminescent reporting ((Van Dyk et al., *Proc. Nat. Acad. Sci. USA* 98:2555-2560 (2001)).

The *luxCDABE* operon has been utilized to create a collection of random gene fusions, comprising 27% of the known or predicted transcriptional units of *E. coli* (Van Dyk et al., *J. Bacteriol.* 183:5496-5505 (2001)). Treatment of *E. coli* cells containing these gene fusions with nalidixic acid, a quinolone, results in selective up-regulation of ten genes. Some of these up-regulated genes are LexA-regulated SOS genes, while others are not generally induced by DNA damage.

Aromatic compounds such as aromatic carboxylic acids are usually toxic to microorganisms. Numerous bacterial strains resistant to aromatic

compounds, however, are known in the prior art (Diaz et al., *Microbiol. Mol. Biol. Rev.* 65:523-569 (2001)). Additionally, a LysR family member from *Acinetobacter*, BenM, is responsive to synergistic induction by benzoic acid, an aromatic carboxylic acid, and muconic acid (Bundy et al., *Proc. Natl. Acad. Sci. USA* 99:7693-7698 (2002)). Benzoic acid alone, however, produces minimal, if any, induction of BenM activity. Even the synergistic response with muconic acid produces only a four-fold increase in BenM activity.

U.S. Patent 5,292,643 issued to Shibano et al. on March 8, 1994 describes genes related to fusaric acid resistance in variety of microorganisms. Specifically, genes capable of decomposing or detoxifying fusaric acid are disclosed. One of the genes postulated to be involved in fusaric acid resistance, *fusB*, shares some homology with the putative efflux transporter (PET) *yhcP* gene (Paulsen et al., *FEMS Microbiol. Lett.* 156:1-8 (1997)). Applicants incorporate by reference the co-owned and concurrently filed application entitled "PET Family of Efflux Proteins", U.S. Patent Application No. 60/440760, which describes new proteins efflux proteins whose expression may alter the expression of carboxylic acids.

The problem to be solved therefore is to discover facile and inexpensive methods of inducible expression of heterologous genes. Applicants have solved the stated problem through the discovery that the promoter elements of the *yhcRQR* operon are responsive to the expression of the *yhcS* regulator (a member of the LysR family of transcriptional regulators) whose expression may be induced by an inexpensive cadre of aromatic carboxylic acid inducers.

SUMMARY OF THE INVENTION

The invention relates to the discovery that the *yhcS* regulator is responsible for the activation of the *yhcRQR* operon promoter and is inducible by aromatic carboxylic acids, compounds that are typically toxic to cells. The expression of heterologous genes, placed under the control of a *yhcRQR* operon promoter may be regulated by the presence of an aromatic carboxylic acid. Accordingly the invention provides a method for the inducible expression of a heterologous nucleic acid molecule comprising:

- a) providing a host cell having a genome comprising:
 - i) a *yhcS* regulator gene responsive to an aromatic carboxylic acid inducer;

ii) a promoter region, responsive to expression of the *yhcS* regulator gene; and

iii) at least one heterologous nucleic acid molecule;

wherein the at least one heterologous nucleic acid molecule is operably linked to the promoter region;

b) contacting the host cell of (a) with an aromatic carboxylic acid inducer wherein the at least one heterologous nucleic acid molecule is expressed.

In a preferred embodiment the invention provides a method for the inducible expression of a heterologous nucleic acid molecule comprising:

a) providing an enteric bacterial host cell having a genome comprising:

i) a *yhcS* regulator gene responsive to an aromatic carboxylic acid inducer;

ii) a promoter region, responsive to expression of the *yhcS* regulator gene; and

iii) at least one heterologous nucleic acid molecule;

wherein the at least one heterologous nucleic acid molecule is operably linked to the promoter region;

b) contacting the host cell of (a) with an aromatic carboxylic acid inducer wherein the at least one heterologous nucleic acid molecule is expressed.

Specific *yhcS* regulator gene useful in the present invention are those that are selected from the group consisting of:

a) an isolated nucleic acid molecule comprising nucleic acid sequence SEQ ID NO:1; and

b) an isolated nucleic acid molecule, which hybridizes to SEQ ID NO:1 after being washed with 0.1×SSC, 0.1% SDS at 65°C and washed with 2×SSC, 0.1% SDS followed by a second wash in 0.2×SSC, 0.1% SDS.

Similarly specific promoter responsive to expression of the *yhcS* regulator gene are those selected from the group consisting of:

a) an isolated nucleic acid molecule comprising nucleic acid sequence SEQ ID NO:3; and

b) an isolated nucleic acid molecule, which hybridizes to SEQ ID NO:3 after being washed with 0.1×SSC, 0.1% SDS at 65°C and washed with 2×SSC, 0.1% SDS followed by a second wash in 0.2×SSC, 0.1% SDS.

In another embodiment the invention provides a host cell comprising:

a) a *yhcS* regulator gene responsive to an aromatic carboxylic acid inducer having a nucleic acid sequence selected from the group consisting of:

i) an isolated nucleic acid molecule comprising nucleic acid sequence SEQ ID NO:1; and

ii) an isolated nucleic acid molecule, which hybridizes to SEQ ID NO:1 after being washed with 0.1×SSC, 0.1% SDS at 65°C and washed with 2×SSC, 0.1% SDS followed by a second wash in 0.2×SSC, 0.1% SDS;

b) a promoter region, responsive to expression of the *yhcS* regulator gene having a nucleic acid sequence selected from the group consisting of:

i) an isolated nucleic acid molecule comprising nucleic acid sequence SEQ ID NO:3; and

ii) an isolated nucleic acid molecule, which hybridizes to SEQ ID NO:3 after being washed with 0.1×SSC, 0.1% SDS at 65°C and washed with 2×SSC, 0.1% SDS followed by a second wash in 0.2×SSC, 0.1% SDS; and

iii) at least one heterologous nucleic acid molecule;

wherein the at least one heterologous nucleic acid molecule is operably linked to the promoter region.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE

DESCRIPTIONS

The invention can be more fully understood from the following detailed description, figures and the accompanying sequence descriptions, which form a part of this application.

Figure 1 shows the kinetics of the *yhcRQP-luxCDABE* response to pHBA in *yhcS*⁺ and *yhcS*⁻ host strains.

The following sequences conform with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide

and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NO:1 is the nucleotide sequence of the *yhcS* regulator.

SEQ ID NO:2 is the amino acid sequence of the YhcS protein.

5 SEQ ID NO:3 is the nucleotide sequence of the promoter region upstream of *yhcQ*.

SEQ ID NO:4 is the nucleotide sequence of the primer Kan-2FP(PCR).

10 SEQ ID NO:5 is the nucleotide sequence of the primer Kan-2RP(PCR).

SEQ ID NO:6 is the nucleotide sequence of the primer Kan-2FP-1.

SEQ ID NO:7 is the nucleotide sequence of the primer Kan-2RP-1.

SEQ ID NO:8 is the nucleotide sequence of the primer YhcS.F.

SEQ ID NO:9 is the nucleotide sequence of the primer YhcS.R.

15 DETAILED DESCRIPTION OF THE INVENTION

There is a need in the field of microbial metabolic engineering for tunable promoters and novel regulatory switches. Advantages of Applicants' system are the very low basal levels of gene expression in the absence of inducer and expression levels that vary with the concentration of inducer up to relatively high expression levels. Furthermore, the cost of many of the inducing molecules is relatively inexpensive.

20 There are numerous uses for Applicants' tunable promoter system comprising the YhcS regulatory protein and the responsive promoter region. This novel promoter/regulator system can be used to control and regulate expression of genes and operons of interest by applying standard molecular biology methods as has been demonstrated herein by the controlled expression of *luxCDABE*.

30 Another possible application is suggested by analogy to other LysR family proteins, which typically have a conformational change upon binding their cognate inducer. YhcS protein likely binds to certain aromatic carboxylic acid molecules and changes conformation such that it activates gene expression upon DNA binding. This conformation change may be useful to sense small molecules in applications such as those found in nano scale systems.

35 The range of molecules to which YhcS responds may be manipulated by application of a variety of protein engineering techniques. Thus, the useful range of molecules of each of the above-mentioned applications could be expanded.

Advantages of this regulator/promoter system include, but are not limited to, basal expression levels that are extremely low; high levels of expression following induction; inexpensive, environmentally friendly inducers; when used in conjunction with other expression systems, an alternative expression system that will allow differential regulation of various genes in a genetically engineered host; and a protein conformation change useful for nanotechnology.

Applicants specifically incorporate the entire content of all cited references in this disclosure.

In the context of this disclosure, a number of terms shall be utilized.

The term "pHBA" is the abbreviation for para-hydroxybenzoic acid, which is also known as para-hydroxybenzoate.

The term "pHCA" is the abbreviation for para-hydroxycinnamic acid, which is also known as para-hydroxycinnamate.

The term "CA" is the abbreviation for cinnamic acid, which is also known as cinnamate

An "isolated nucleic acid molecule" refers to a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid molecule in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference) (hereinafter "Sambrook"). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6×SSC, 0.5% SDS at room temperature for 15 min,

then repeated with 2×SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2×SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2×SSC, 0.5% SDS is increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1×SSC, 0.1% SDS at 65°C. An additional set of stringent conditions include hybridization at 0.1×SSC, 0.1% SDS, 65°C and washed with 2×SSC, 0.1% SDS followed by a second wash in 0.2×SSC, 0.1% SDS, for example.

Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (Sambrook *supra*). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (Sambrook *supra*). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably, a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

A "substantial portion" refers to an amino acid or nucleotide sequence which comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al., *J. Mol. Biol.* 215:403-

410 (1993). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific
5 oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., in situ hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to
10 obtain a particular nucleic acid molecule comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid molecule comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide
15 sequences encoding one or more particular bacterial proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for the purpose known to those skilled in the art. Accordingly; the instant invention comprises the complete sequences as reported in the accompanying
20 Sequence Listing, as well as substantial portions of those sequences as defined above.

The term "complementary" describes the relationship between nucleotide bases that are capable to hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine
25 and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid molecules that are complementary to the complete sequences as reported in the accompanying Sequence Listing as well as those substantially similar nucleic acid sequences.

30 The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as
35 determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing*:

Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data*, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press, New York (1987); and *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG Pileup program found in the GCG program package, using the Needleman and Wunsch algorithm with their standard default values of gap creation penalty=12 and gap extension penalty=4 (Devereux et al., *Nucleic Acids Res.* 12:387-395 (1984)), BLASTP, BLASTN, and FASTA (Pearson et al, *Proc. Natl. Acad. Sci. USA* 85:2444-2448 (1988)). The BLASTX program is publicly available from NCBI and other sources (BLAST Manual, Altschul et al., *Natl. Cent. Biotechnol. Inf., Natl. Library Med. (NCBI NLM) NIH, Bethesda, Md.* 20894; Altschul et al., *J Mol. Biol.* 215:403-410 (1990); Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)). Another preferred method to determine percent identity is by the method of DNASTAR protein alignment protocol using the Jotun-Hein algorithm (Hein et al., *Meth. Enzymol.* 183:626-645 (1990)). Default parameters for the Jotun-Hein method for alignments are: for multiple alignments, gap penalty=11, gap length penalty=3; for pairwise alignments ktuple=6. As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed

either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy-terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

The term "percent homology" refers to the extent of amino acid sequence identity between polypeptides. When a first amino acid sequence is identical to a second amino acid sequence, then the first and second amino acid sequences exhibit 100% homology. The homology between any two polypeptides is a direct function of the total number of matching amino acids at a given position in either sequence, e.g., if half of the total number of amino acids in either of the two sequences is the same then the two sequences are said to exhibit 50% homology.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon

usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

5 "Gene" refers to a nucleic acid molecule that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in
10 nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

"Genome" refers to the entire genetic information contained within
15 an organism (e.g., chromosome, plasmid, plastid, or mitochondrial DNA). "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a
20 non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure. "Structural gene" refers to a gene that codes for the amino acid sequence of a protein or for a ribosomal RNA or transfer RNA. An "operon" refers to a controllable unit of transcription consisting of a number of structural
25 genes transcribed together.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and
30 which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the
35 expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even

comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters which
5 cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

10 "Heterologous" as used in the context of gene expression relates to that which is "foreign" to a particular environment. Thus, a "heterologous gene" or "heterologous nucleic acid molecule" means a nucleic acid molecule that is foreign, or non-native to a particular host or genome. A "heterologous protein" is a protein that is foreign to a host cell and is
15 typically encoded by a heterologous gene. Heterologous nucleic acids of the invention are typically expressed under the control of regulated promoters in an inducible fashion.

The term "regulator" refers to a protein whose primary function is to control the rate of expression of regulated genes. Regulation of gene
20 expression can be by positive activation or by repression. A "regulatory gene" refers to a gene that encodes a regulator. Within the context of the present invention a typical regulator gene is the *yhcS* gene

"Host cell" refers to a cell into which has been introduced (e.g., transformed or transfected) an exogenous polynucleotide sequence, i.e. a
25 heterologous nucleic acid molecule. Host cells are typically prokaryotic cells such as bacteria, e.g., *E. coli*, and may be eukaryotic cells such as yeast, insect, amphibian, green plant, or mammalian cells, where the relevant regulator genes exist.

"Inducer" refers to a small molecule that initiates transcription, or
30 increases the rate of transcription, of a desired gene. Within the context of the present invention typical inducers are aromatic carboxylic acids having the ability to activate a regulator gene.

"Reporter gene" refers to a gene that encodes an easily assayed product (e.g. *luxCDABE*, *bgaB*, *cat*, *dsRed*, *galK*, *gfp*, *lacZ*, *luc*, *luxAB*,
35 *nptII*, *phoA*, *uidA*, or *xylE*). Typically reporters are coupled to the promoter and/or regulator sequence of another gene and transfected into a host cell. The reporter gene can then be used to see which factors activate response elements in the upstream region of the gene of interest.

"Translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader
5 sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner et al., *Mol. Biotechnol.* 3:225 (1995)).

"3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition
10 sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., *Plant Cell*
15 1:671-680 (1989).

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived
20 from post transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that
25 includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific
30 gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet and has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid
35 sequences on a single nucleic acid molecule so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it affects the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the

promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

5 The term "expression" refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid molecule of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or
10 nontransformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020).

15 "Transformation" refers to the transfer of a nucleic acid molecule into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms.

20 The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes that are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have
25 been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate
30 transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

35 "PCR" or "polymerase chain reaction" is a technique used for the amplification of specific DNA segments (U.S. Patent Nos. 4,683,195 and 4,800,159).

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, J., Fritsch,

E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) (hereinafter "Sambrook"); and by Silhavy, T. J., Bannan, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Cold Press Spring Harbor, NY (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, published by Greene Publishing Assoc. and Wiley-Interscience (1987).

The present invention relates to the discovery that aromatic carboxylic acids induce the up-regulation or expression of the regulator gene *yhcS* which in turn is responsible for the activation of the promoter region driving expression of the *yhcRQP* operon. Expression of heterologous nucleic acid molecules which are operably to these promoters may therefore be inducibly regulated by the presence or absence of inexpensive aromatic carboxylic acids in the medium.

15 Regulator Systems

The *yhcS* regulator and the corresponding *yhcRQR* operon is known and is common in a variety of enteric bacteria such as *Escherichia* (Hayashi et al., "Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12", *DNA Res.* 8 (1), 11-22 (2001)); *Yersinia* (Parkhill et al., "Genome sequence of *Yersinia pestis*, the causative agent of plague", *Nature* 413 (6855), 523-527 (2001)); *Shigella* (Wei et al., "Complete Genome Sequence and Comparative Genomics of *Shigella flexneri* Serotype 2a Strain 2457T", *Infect. Immun.* 71 (5), 2775-2786 (2003)); and *Salmonella* (McClelland et al., "Complete genome sequence of *Salmonella enterica* serovar *Typhimurium* LT2", *Nature* 413 (6858), 852-856 (2001)). It will be appreciated by one of skill in the art that those organisms having homologs to the present regulators will be expected to function in heterologous gene expression in similar ways.

Those cells having existing homologous regulator systems may be used in the present invention for the expression of heterologous DNA simply by the insertion of the DNA to be expressed in the appropriate position in the genome and in the correct orientation for expression. Thus a *Salmonella* or *Shigella* strain, as described above, may be used in this fashion. Host cells suitable for use in the present invention will include but are not limited to *Escherichia*, *Salmonella*, *Bacillus*, *Acinetobacter*, *Streptomyces*, *Methylobacter*, *Rhodococcus*, *Corynebacterium*, *Pseudomonas*, *Rhodobacter*, and *Synechocystis*.

Particularly suitable in the present invention are members of the enteric class of bacteria. Enteric bacteria are members of the family *Enterobacteriaceae* and include such members as *Escherichia*, *Salmonella*, and *Shigella*. They are gram-negative straight rods, 0.3-1.0 X 1.0-6.0 mm, motile by peritrichous flagella (except for *Tatumella*) or nonmotile. They grow in the presence and absence of oxygen and grow well on peptone, meat extract, and (usually) MacConkey's media. Some grow on D-glucose as the sole source of carbon, whereas others require vitamins and/or mineral(s). They are chemoorganotrophic with respiratory and fermentative metabolism but are not halophilic. Acid and often visible gas is produced during fermentation of D-glucose, other carbohydrates, and polyhydroxyl alcohols. They are oxidase negative and, with the exception of *Shigella dysenteriae* 0 group 1 and *Xenorhabdus nematophilus*, catalase positive. Nitrate is reduced to nitrite (except by some strains of *Erwinia* and *Yersinia*). The G + C content of DNA is 38-60 mol% (T_m , Bd). DNAs from species within most genera are at least 20% related to one another and to *Escherichia coli*, the type species of the family. Notable exceptions are species of *Yersinia*, *Proteus*, *Providencia*, *Hafnia* and *Edwardsiella*, whose DNAs are 10-20% related to those of species from other genera. Except for *Erwinia chrysanthemi*, all species tested contain the enterobacterial common antigen (Bergey's Manual of Systematic Bacteriology, D. H. Bergey et al., Baltimore: Williams and Wilkins, 1984).

It is clear that host cells comprising the present regulator systems are suitable for use in the invention. However, where it is desired to find new strains having the present regulator systems, or to identify new regulator genes having greater functionality in non-native host cells, it will be possible to use the sequence information provided in the literature and in this disclosure to identify and isolate such homologs.

Isolation of Homologs

A specific *yhcS* regulator has been identified in the *E. coli* genome and has the nucleic acid sequence as set forth in SEQ ID NO:1. The promoter region of the *yhcRQP* operon, responsive to the expression of this *yhcS* regulator has the nucleic acid sequence as set forth in SEQ ID NO:3. It will be apparent to the skilled artisan that homologs to the *E. coli* sequences or others cited in the literature may easily be identified based on current practices in molecular biology, and such homologs will be equally applicable and useful in the present invention. For example, one

of skill in the art may use the nucleic acid molecules of the instant invention to isolate cDNAs and genes encoding a homologous YhcS protein from the same or other bacterium species. Isolation of homologous genes using sequence-dependent protocols is well known in the art.

5 Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., PCR or ligase chain reaction).

10 For example, *yhcS* gene, either as cDNA or genomic DNA, could be isolated directly by using all or a portion of the instant nucleic acid molecule as DNA hybridization probes to screen libraries from any desired bacterium employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant *yhcS* gene sequence can be designed and synthesized by methods known in the art
15 (Sambrook *supra*). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primers DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available in vitro transcription systems. In addition, specific primers can be designed and used to amplify a part of or full-
20 length of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

25 In addition, two short segments of the instant nucleic acid molecule may be used in polymerase chain reaction protocols to amplify longer nucleic acid molecules encoding homologous *yhcS* genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid molecules wherein the sequence of one primer is derived from the instant nucleic acid molecules, and the sequence of the
30 other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., *Proc. Natl. Acad. Sci. USA* 85:8998-9002 (1988)) to
35 generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (Invitrogen,

Carlsbad, CA), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., *Proc. Natl. Acad. Sci. USA* 86:5673-5677 (1989); Loh et al., *Science* 243:217-220 (1989)). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman et al., *Techniques* 1:165 (1989)).

Alternatively the *yhcS* sequences may be employed as an hybridization reagent for the identification of homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the gene or gene fragment of interest, and a specific hybridization method. Probes are typically single stranded nucleic acid sequences which are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. The probe length can vary from 5 bases to tens of thousands of bases, and will depend upon the specific test to be done. Typically a probe length of about 15 bases to about 30 bases is suitable. Only part of the probe molecule need be complementary to the nucleic acid sequence to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions which will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration the shorter the hybridization incubation time needed. Optionally a chaotropic agent may be added. The chaotropic agent stabilizes nucleic acids by inhibiting nuclease activity. Furthermore, the chaotropic agent allows sensitive and stringent hybridization of short oligonucleotide probes at room temperature (Van Ness and Chen, *Nucl. Acids Res.* 19:5143-5151(1991)). Suitable chaotropic agents include guanidinium chloride, guanidinium thiocyanate, sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium tetrachloroacetate, potassium iodide, and cesium trifluoroacetate,

among others. Typically, the chaotropic agent will be present at a final concentration of about 3M. If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).

Various hybridization solutions can be employed. Typically, these
5 comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1M sodium chloride, about 0.05 to 0.1M buffers, such as sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9), about 0.05 to 0.2% detergent, such as sodium dodecylsulfate, or between
10 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kilodaltons), polyvinylpyrrolidone (about 250-500 kdal), and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA, e.g., calf thymus or salmon sperm DNA, or yeast RNA, and optionally from
15 about 0.5 to 2% wt./vol. glycine. Other additives may also be included, such as volume exclusion agents which include a variety of polar water-soluble or swellable agents, such as polyethylene glycol, anionic polymers such as polyacrylate or polymethylacrylate, and anionic saccharidic polymers, such as dextran sulfate.

20 Nucleic acid hybridization is adaptable to a variety of assay formats. One of the most suitable is the sandwich assay format. The sandwich assay is particularly adaptable to hybridization under non-denaturing conditions. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it
25 immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

Heterologous DNA Expression

Once a host cell comprising a *yhcS* regulator and the responsive *yhcRQP* operon has been identified or constructed it will be necessary to
30 insert the heterologous nucleic acid molecule into the genome in the position that will be appropriate for expression. Methods for the transformation of microbial cells and integration of DNA into a genome are common and well known in the art (see Sambrook *supra*). Typically a transformation vector is constructed for this purpose containing the
35 essential elements for transformation and DNA integration. Typically the vector or cassette contains sequences directing transcription and translation of the relevant gene, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable

vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. It is most preferred when both control regions are derived from genes homologous to the transformed host cell, although
5 it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.

Heterologous nucleic acid molecules suitable for expression by the methods of the invention are virtually unlimited. In some instances the heterologous nucleic acid molecule may not encode a protein and be
10 expressed for the purpose of controlling or suppressing (as in antisense orientation for example) other genetic elements in the genome. More commonly, the foreign DNA will encode a protein and typically an enzyme that is part of a pathway. It will be appreciated that a single DNA fragment may be expressed by the present method or several linked fragments
15 comprising all or a part of an enzymatic pathway. It is therefore within the scope of the present invention to express heterologous nucleic acid molecules encoding at least one protein wherein the at least one protein is part of an enzymatic biosynthetic pathway producing a product selected from the group consisting of isoprenoids; terpenoids; tetrapyrroles,
20 polyketides, vitamins, amino acids, fatty acids, proteins, nucleic acids, carbohydrates, antimicrobial agents, and anticancer agents.

In some instances it will be useful to monitor the expression or activation of the regulator systems through the use of a reporter. Reporters are common and well known in the art and a non-inclusive list of
25 those suitable in the present invention are luxCDABE, bgaB, cat, dsRed, galK, gfp, lacZ, luc, luxAB, nptII, phoA, uidA, and xylE.

One of the key advantages of the present invention is the ability to control the expression of the heterologous DNA by the action of an inducer. Applicant's discovery that the present regulator genes are
30 responsive to aromatic carboxylic acids is fortuitous in that these compounds are inexpensive and easily obtained. Any aromatic carboxylic acid will have utility in the present method where para-hydroxybenzoic acid, para-hydroxycinnamic acid, cinnamic acid, salicylic acid, benzoic acid, and 1-napthoic acid are preferred.

35 Description of the Preferred Embodiments

As described in the following examples, during the course of LuxArray and DNA array analyses, Applicants discovered that expression of an *E. coli* operon, *yhcRQP*, was highly induced by treatment of *E. coli*

cells with aromatic carboxylic acids, such a para-hydroxybenzoic acid (pHBA), para-hydroxycinnamic acid (pHCA), and cinnamic acid (CA). A *luxCDABE* gene fusion with the promoter region of this operon controlling the bioluminescent reporter was used to characterize expression of this operon. In the absence of inducing molecules the expression level was very low, just slightly above background. However, when the cell was treated with inducer, the expression level was dramatically elevated, where expression increased nearly 1000-fold. This represents much higher expression than found in typical LysR family members. Furthermore, the expression level was dependent on the concentration of inducer added.

The gene immediately upstream of the *yhcRQP* operon at open reading frame b3243, *yhcS*, is a putative member of the LysR family of transcription regulators. Using a transposon insertion mutation in *yhcS*, expression of the *yhcRQP-luxCDABE* gene fusion was no longer induced by aromatic carboxylic acids. Thus, it was shown that *yhcS* encodes a positive-acting transcriptional regulator responsible for the dramatic, tunable changes in gene expression of the *yhcRQP* operon.

This promoter/regulator system can be used to control and regulate expression of other genes and operons of interest by applying standard molecular biology methods. Furthermore, by analogy to other LysR family proteins, the YhcS protein likely binds to these aromatic carboxylic acid molecules and changes conformation such that it activates gene expression upon DNA binding. This conformation change may be useful to sense small molecules in nano-scale systems.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees in Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usage and conditions.

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook,

J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY (1989); by T. J. Silhavy, M. L. Bannan, and L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987).

Standard genetic methods for transduction used in the Examples are well known in the art and are described by Miller, J. H., *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1972).

The meaning of abbreviations is as follows: "kb" means kilobase(s), "hr" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "ml" means milliliter(s), " μ l" means microliter(s), "nl" means nanoliter(s), " μ g" means microgram(s), "ng" means nanogram(s), "mM" means millimolar, " μ M" means micromolar, "nm" means nanometer(s), "OD₆₀₀" means the optical density measured at a wavelength of 600 nm, "RLU" means relative light units.

Media and Culture Conditions:

Materials and methods suitable for the maintenance and growth of bacterial cultures were found in *Experiments in Molecular Genetics* (Jeffrey H. Miller), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1972); *Manual of Methods for General Bacteriology* (Phillip Gerhardt, R.G.E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), pp. 210-213, American Society for Microbiology, Washington, DC (1981); or Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA. All reagents and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), BD Diagnostic Systems (Sparks, MD), Invitrogen Corp. (Carlsbad, CA), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

LB medium contains the following per liter of medium: Bacto-tryptone (10 g), Bacto-yeast extract (5 g), and NaCl (10 g).

Vogel-Bonner medium contains the following per liter: 0.2 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 2 g citric acid $\cdot \text{H}_2\text{O}$, 10 g K_2HPO_4 and 3.5 g $\text{NaNH}_4\text{HPO}_4 \cdot 4\text{H}_2\text{O}$.

Minimal M9 medium contains the following per liter of medium:
Na₂HPO₄ (6 g), KH₂PO₄ (3 g), NaCl (0.5 g), and NH₄Cl (1 g).

Above media were autoclaved for sterilization then 10 ml of 0.01 M CaCl₂ and 1 ml of 1 M MgSO₄·7H₂O were added to M9 medium. Vitamin B1 (thiamin) was added at 0.0001% to both Vogel-Bonner and M9 media. Carbon source and other nutrients and supplements were added as mentioned in the Examples. All additions were pre-sterilized before they were added to the media.

Molecular Biology Techniques:

Restriction enzyme digestions, ligations, transformations, and methods for agarose gel electrophoresis were performed as described in Sambrook *supra*. Polymerase Chain Reactions (PCR) techniques were found in White, B., *PCR Protocols: Current Methods and Applications*, Volume 15 (1993) Humana Press Inc, Totowa, NJ.

Example 1: Gene Expression Profiling of para-hydroxybenzoate-treated *E. coli* Cells

The alterations in the *E. coli* gene expression profile upon exposure to pHBA were examined using DNA microarray technology. *E. coli* strain DE112 (Van Dyk et al. *Appl. Environ. Microbiol.* 60:1414-1420 (1994)) was grown in Vogel-Bonner minimal medium with glucose as a carbon source to an OD₆₀₀ of 0.2. At this point the culture was split in two flasks and pHBA in the acid form was added to one flask from a stock solution in ethanol to achieve a final pHBA concentration of 25 mM. The pH of the medium in the flask with pHBA was lowered by an unmeasured amount. An equivalent volume of ethanol without pHBA was added to the other flask. Approximately 58% growth inhibition resulted from the pHBA treatment under these conditions. Cells were harvested from the control and treated flasks at 30 and 60 minutes after pHBA addition. RNA isolation, array hybridization, and data analysis were done as previously described (Wei et al, *J. Bacteriol.* 183:545-556 (2001), Smulski et al. *J. Bacteriol.* 183:3353-3364 (2001)). Among the genes that were highly induced by this treatment at the 60 minute time point were the *yhcR*, *yhcQ*, and *yhcP* genes (Table 1). The experiment was repeated, and again, these three genes were highly upregulated (Table 2). These three genes are predicted to be transcribed as an operon. The reproducible observed co-regulation in response to pHBA treatment is consistent with this

prediction. Thus, these three genes will be referred to as the *yhcRQP* operon.

Table 1: Upregulation of genes in the *yhcRQP* operon after 60 minutes pHBA treatment, experiment 1

Gene	Blattner No.*	Signal in untreated	Signal in pHBA treated	Ratio (treated/untreated)
<i>yhcR</i>	b3242	0.0190	1.37	72.3
<i>yhcQ</i>	b3241	0.393	1.61	4.09
<i>yhcP</i>	b3240	1.17	7.52	6.45

Table 2: Upregulation of genes in the *yhcRQP* operon after 60 minutes pHBA treatment, experiment 2

Gene	Blattner No.*	Signal in untreated	Signal in pHBA treated	Ratio (treated/untreated)
<i>yhcR</i>	b3242	0.687	7.18	10.5
<i>yhcQ</i>	b3241	0.270	6.02	22.3
<i>yhcP</i>	b3240	0.994	12.0	12.1

*Blattner et al. *Science* 277:1453-1462 (1997).

Example 2: LuxArray analysis of para-hydroxycinnamic acid and cinnamic acid treated *E. coli*

Gene expression profiles were done using LuxArray version 1.04, which has been fully described (Van Dyk et al. *J. Bacteriol.* 183:5496-5505 (2001), Gonye et al. U.S. Patent Application Publication 20030219736).

This method utilizes a set of bioluminescent gene fusions to 1/3 of *E. coli* transcriptional units in a *tolC*⁻ host strain that is hypersensitive to many compounds, including pHCA. Sublethal concentrations of pHCA, 10 mM and 5 mM, and CA, 8 mM and 4 mM, at which stress and other responses can be detected using *luxCDABE* reporter gene fusions, were used. Each of these treatments yielded nearly identical expression patterns, suggesting a similar cellular response to these two aromatic compounds. A predominant feature observed in both these profiles was that one gene fusion, lux-a.pk035.c7, was the several hundred-fold upregulated. The lux-a.pk035.c7 gene fusion contains an *E. coli* chromosomal segment between nucleotides 3385829 and 3386761 according to the *E. coli* genomic sequence, which contains the promoter region of the putative *yhcRQP* operon and the entire *yhcR* gene and the 5' end of the *yhcQ* gene. This chromosomal segment is joined to the *luxCDABE* gene fusion

in the parental plasmid, pDEW201 (Gonye et al. U.S. Patent Application Publication 20030219736) thus forming junction between *yhcQ* gene and *luxC*. Accordingly, this gene fusion will report on expression of the *yhcQ* gene and any other genes cotranscribed with it. As detailed in Example 1, it is likely that *yhcR*, *yhcQ*, and *yhcP* form an operon, thus this gene fusion is referred to as a *yhcRQP-luxCDABE* gene fusion. The strain containing this gene fusion was given the name DPD2411, and the plasmid within this strain that contains *yhcRQP-luxCDABE* gene fusion was called pDEW655.

Table 3 shows the bioluminescent response of this reporter gene fusion to pHCA, CA, and ethanol. As part of a larger LuxArray experiment, two independent, actively growing, cultures carrying the *yhcRQP-luxCDABE* gene fusion in LB medium were each split three ways at time zero. Two aliquots were treated with different concentrations of each chemical and the third aliquot was the untreated control. The normalized bioluminescent signal from each of two replicas in the LuxArray is shown for the measurements at four time points (in minutes).

The dramatic upregulation of expression in response to pHCA and CA treatments at each of the time points other than the initial, zero time point is clear. In contrast, ethanol treatment does not induce increased bioluminescence.

Table 3: Responses of the *yhcRQP-luxCDABE* gene fusion to pHCA, CA, or ethanol

Treat- ment	Replica 1 normalized RLU				Replica 2 normalized RLU			
	0 min	45 min	90 min	135 min	0 min	45 min	90 min	135 min
0 mM pHCA	0.061	0.071	0.067	0.039	0.043	0.04	0.039	0.022
5 mM pHCA	0.056	10.695	14.276	13.533	0.043	9.325	14.754	10.587
10 mM pHCA	0.056	9.094	14.314	15.283	0.034	7.31	11.164	12.867
0 mM CA	0.03	0.034	0.024	0.013	0.08	0.052	0.032	0.015
4 mM CA	0.026	4.092	5.845	5.394	0.062	2.928	4.951	3.678

8 mM CA	0.022	4.937	8.379	6.981	0.054	3.405	6.193	5.746
0% ethanol	0.021	0.023	0.017	0.009	0.045	0.028	0.018	0.008
3% ethanol	0.024	0.024	0.022	0.012	0.038	0.029	0.026	0.014
5% ethanol	0.022	0.014	0.022	0.015	0.037	0.017	0.017	0.014

Example 3: Regulation of *yhcRQP* expression by YhcS

5 The *yhcS* gene of *E. coli* encodes an uncharacterized member of the LysR family of positive acting regulatory molecules. This gene is located immediately adjacent to the *yhcRQP* operon that was found to be upregulated by pHBA treatment in DNA array experiments and by pHCA and CA treatments in LuxArray experiments. The possibility that YhcS
10 controls expression of *yhcRQP* was tested using a *yhcS* null mutation.

Such a mutation was found in an *E. coli* library of transposon insertion mutations constructed using the transposome system based on the Tn5 transposon (Epicentre, Madison, WI). A transposome is a protein-DNA complex composed of the EZ::TN<Kan-1> transposon and the
15 EZ::TN transposase. The EZ::TN transposase is bound to the ends of the transposon, which facilitates the formation of a stable synaptic complex. The transposome requires Mg^{+2} to initiate the insertion of the EZ::TN<Kan-1> transposon into target DNA. The cellular levels of Mg^{+2} are sufficient to activate the transposome. Thus, the electroporation of the transposome
20 into cells permits the in vivo insertion of the EZ::TN<Kan-1> transposon into bacterial genomes.

The EZ::TN<Kan-1> transposome was electroporated into electroporation competent *E. coli* strain DH5 α E cells (Invitrogen, Carlsbad, CA). Following electroporation, the cells were grown in SOC medium
25 (Invitrogen) for one hour at 37°C with aeration. Subsequently, the cells were plated onto LB agar plates containing kanamycin (50 μ g/ml) (LB+Kan) and incubated overnight at 37°C. Individual colonies were inoculated into 96-well microtiter plates containing 150 μ l of LB+Kan and incubated overnight at 37°C.

“Single Primer PCR” was used to determine the identity of each *E. coli* transposon mutation. Using a single DNA primer that was complementary to one end of the EZ::TN<Kan-1> transposon, PCR products were generated. Subsequently, a second DNA primer (located
5 internal and adjacent to the PCR primer) was used to sequence the PCR products. The DNA primer used in the PCR reaction was either Kan-2FP(PCR) (SEQ ID NO:4) or Kan-2RP(PCR) (SEQ ID NO:5) and the DNA primer used for DNA sequencing was either Kan-2FP(PCR) (SEQ ID NO:4) or Kan-2RP(PCR) (SEQ ID NO:5), respectively. The PCR reaction
10 conditions were the following: (1) 94°C, 15 minutes (2) 20 cycles - 94°C, 30 seconds; 60°C, 30 seconds; 72°C, 3 minutes (3) 30 cycles – 94°C, 30 seconds; 40°C, 30 seconds; 72°C, 2 minutes (4) 30 cycles – 94°C, 30 seconds; 60°C, 30 seconds; 72°C, 2 minutes (5) 72°C, 7 minutes. The PCR reactions were prepared for DNA sequencing using the QIAquick
15 PCR Purification Kit (Qiagen, Valencia, CA).

The *yhcS* transposon mutant was identified using PCR amplification primer Kan-2FP(PCR) (SEQ ID NO:4) and DNA sequencing primer Kan-2FP-1 (SEQ ID NO:6). The transposon mutation was confirmed using
gene-specific primers: YhcS.F (SEQ ID NO:8) and YhcS.R (SEQ ID NO:9)
20 and transposon-specific primers Kan-2FP-1 (SEQ ID NO:6) and Kan-2RP-1 (SEQ ID NO:7).

The size of the *yhcS* gene is ~929 base pairs. The transposon insertion site within the *yhcS* gene is ~ 330 base pairs away from the 5' end of *yhcS*. A PCR reaction done with the YhcS.F and Kan-2RP-1
25 primers yielded a PCR fragment ~550 base pairs and PCR primers YhcS.R and Kan-2FP-1 yielded a PCR product <400 base pairs in size.

E. coli strain DPD2410 is the DH5 α E derived strain containing the *yhcS*::TN<Kan> mutation. Strains DH5 α E and DPD2410 were transformed with pDEW655 to generate *E. coli* strains DPD2413 and
30 DPD2415, respectively. A single colony of each of these two strains from an LB plate containing 150 μ g/ml Ampicillin was used to inoculate 200 μ l LB medium in wells of a 96 well, white microplate (Microlite, Dynex Technologies, Chantilly, VA). The plate was incubated for 90 minutes at 37°C; then 50 μ l of the cultures was added to 50 μ l of LB medium or to 50
35 μ l of LB medium containing pHBA in the acid form, which had been added from a stock solution in ethanol. The final concentration of pHBA was 5 mM and the final concentration of ethanol was 0.25%. The pH of the medium with pHBA was lowered by an unmeasured amount. The

bioluminescence was quantitated with a Luminoskan Ascent microplate luminometer (Thermo Labsystems, Franklin, MA). The results of this study are presented in Figure 1, which is a plot of the bioluminescence intensity in relative light units (RLU) versus time in minutes. In the Figure, addition of pHBA was made at time zero. Solid lines are the response in the *yhcS*⁺ strain, DPD2413. Dotted lines are the response in the *yhcS*⁻ strain, DPD2415. Circles represent pHBA treated cultures and triangles represent untreated cultures. Figure 1 clearly shows that pHBA induced rapid and dramatic upregulation of the *yhcRQP-luxCDABE* expression in the *yhcS*⁺ host strain, but that this upregulation was essentially abolished in the *yhcS::TN<Kan>* host strain.

A derivative of *E. coli* strain MG1655 (obtained from Prof. Douglas Berg, Washington University School of Medicine, St. Louis, MI) with the *yhcS::TN<Kan>* mutation was made by P1clr100Cm mediated transduction using phage grown on strain DPD2410 as a donor and selection for kanamycin resistance. The presence of the *yhcS::TN<Kan>* mutation in one of the resultant transductants, named DPD2433, was confirmed by PCR amplification. Plasmid pDEW655 was moved to *E. coli* strains MG1655 and DPD2433 by transformation, selecting for Ampicillin resistance to generate strains DPD2436 and DPD2437, respectively. The bioluminescent response of these two strains to pHBA was tested. Aliquots (50 μ l) of actively growing cultures at 37°C in LB medium that had been previously diluted, and from overnight cultures in LB medium with 150 μ g/ml Ampicillin were added to 50 μ l of LB medium at pH 7.0 containing pHBA as the sodium salt form. Several concentrations of pHBA were tested. Table 4 below shows the response in these two host strains at thirty minutes after cells were added to pHBA containing medium.

Table 4: Bioluminescence response of the *yhcRQP-luxCDABE* gene fusion

[pHBA], mM	RLU		Ratio treated/control	
	<i>yhcS</i> ⁺	<i>yhcS</i> ⁻	<i>yhcS</i> ⁺	<i>yhcS</i> ⁻
100	0.437	0.045	0.693	0.055
50	91.7	0.614	145	0.753
25	66.6	1.59	106	1.95
12.5	30.8	1.82	48.8	2.23
6.2	16.2	1.42	25.7	1.75

3.1	10.2	1.16	16.2	1.42
1.6	6.72	1.02	10.6	1.25
0	0.631	0.815	1	1

The *yhcS*::TN<Kan> mutation almost completely eliminated the upregulation of expression induced by pHBA treatment at all concentrations tested. Also note that in the *yhcS*⁺ strain, the level of gene expression as quantitated by the degree of bioluminescence varied with the concentration of pHBA added. Thus, the amount of inducer added can be used to tune the expression level from this promoter.

Overall, in two different *E. coli* host strains, a functional YhcS was required for upregulation of *yhcRQP* expression in response to pHBA addition. These results prove that YhcS is a positive acting factor for upregulation of transcription of the *yhcRQP* operon.

Example 4: Structure activity relationships for YhcS activation

Further characterization of the signals that trigger activation of YhcS, was done using the *yhcRQP-luxCDABE* gene fusion containing *E. coli* strains DPD2411 or DPD2436. Table 5 shows the results of bioluminescence activation tests done with cells in LB medium at pH 7.0, as described in Example 3. Several weak, aromatic acid molecules in addition to those shown in the Examples above activated expression. Thus, the known inducing molecules comprise pHBA, pHCA, CA, salicylate, benzoate, and 1-napthoate.

Table 5: Upregulation of *yhcRQP-luxCDABE* expression by aromatic carboxylic acids

Experiment code*	<i>E. coli</i> strain	Compound	Concentration of maximum response	treated RLU	control RLU	Ratio
A	DPD2411	1-napthoate	2.5 mM	29.662	0.2244	132
A	DPD2411	Sodium pHBA	25 mM	69.678	0.231	302
B	DPD2436	Sodium benzoate	6.2 mM	3.759	0.2898	13
C	DPD2436	Sodium benzoate	12.5 mM	8.093	0.6792	12

C	DPD2436	Sodium pHBA	50 mM	91.658	0.6311	145
D	DPD2436	Sodium salicylate	6.2 mM	14.703	0.1908	77
D	DPD2436	Sodium pHBA	50 mM	48.643	0.196	248

* Experiments with the same code letter were done simultaneously using one split *E. coli* culture.

Compounds tested that did not induce expression were defined as those for which there resulted less than 3-fold increase in light production from *E. coli* strains containing pDEW655. Compounds unrelated in structure to the known inducing molecules did not induce expression. Those tested were acetate, propionate, ethanol, limonene, NaCl, polymyxin sulfate, benzalkonium chloride, gramicidin S, and SDS. In addition, several compounds related in structure to the inducing molecules were not inducers, including methyl paraben, p-hydroxystyrene, 2-biphenylcarboxylate, and L-tyrosine. Thus, the requirement for the carboxylate moiety was demonstrated by the lack of response to methyl paraben, the methyl ester of pHBA, and to p-hydroxystyrene, a molecule related to pHBA but lacking the carboxylate group. The requirement for an aromatic ring was demonstrated by the lack of response to non-aromatic carboxylic acids, acetate and propionate.

The response of this regulatory system is specific for certain aromatic carboxylic acids. This class of molecules includes compounds that are environmentally friendly and relatively inexpensive, such as sodium benzoate.

Example 5: Internal acidification is not the signal that activates YhcS

All characterized activators of YhcS are weak acids such as pHBA and pHCA. Thus, the inducing condition could potentially be either acidification of the cytoplasm or presence of the conjugate molecule. The fact that non-aromatic weak acids propionate and acetate, which are known to cause cytoplasmic acidification, did not induce expression of the *yhcRQP-luxCDABE* gene fusion suggested that cytoplasmic acidification was not the inducing signal. This conclusion was confirmed by experiments comparing upregulation of *yhcRQP-luxCDABE* mediated by YhcS to that mediated by other well-known acidification responsive

regulatory circuits. Three *E. coli* strains, each in the same host strain but carrying different, plasmid-borne, promoter-*luxCDABE* fusions were used. Strain DPD2411 contains a *yhcRQP-luxCDABE* gene fusion as described in Example 2. Strain DPD2084 contains a *yciG-luxCDABE* gene fusion that has been previously described. Strain DPD3282 contains a *lysU-luxCDABE* gene fusion that was part of the LuxA collection of gene fusions, described in Example 2. The plasmid in this strain, pDEW558, contains an *E. coli* chromosomal segment between nucleotides 4350990 and 4353107 according to the *E. coli* genomic sequence; the orientation of the chromosomal segment is such that the *lysU* promoter region controls expression of *luxCDABE*. Each of these three strains was grown overnight in Vogel-Bonner minimal medium with 0.4% glucose as the carbon source and supplemented with L-proline, L-lysine, uracil and 25 μ g/ml Ampicillin. The overnight cultures were diluted into the same medium except lacking Ampicillin and incubated at 37°C until in mid-exponential growth. Aliquots (50 μ l) of these actively growing cultures were added to 50 μ l of the same medium at pH 7.0 without Ampicillin but containing various concentrations of sodium acetate or sodium salicylate in the wells of a 96 well, white microplate (Microlite, Dynex Technologies). Immediately after adding the cell culture, the bioluminescence was quantitated in a microplate luminometer in the kinetic mode. Table 6 shows the results at 100 minutes after acetate addition or salicylate addition. Treatment of *E. coli* with acetate did not activate expression of *yhcRQP-luxCDABE*, but did activate the other two acid responsive regulatory circuits. Conversely, addition of sodium salicylate upregulated expression of *yhcRQP-luxCDABE*, but did not increase expression of the other two acid responsive gene fusions at the concentrations tested. Thus, it can be concluded that YhcS is not responding to acidification signals, but rather is responding to the presence of the aromatic molecules.

Table 6: Comparison of acid inducible gene fusions and *yhcRQP-luxCDABE* responses to acetate and salicylate

Gene fusion	RLU at 100 minutes					
	0 mM acetate	80 mM acetate	160 mM acetate	0 mM salicylate	0.6 mM salicylate	5.0 mM salicylate
<i>yhcRQP-</i>	0.13	0.14	0.08	0.32	7.9	26.0

<i>luxCDABE</i>						
<i>yciG-luxCDABE</i>	0.19	0.35	0.80	0.18	0.19	0.13
<i>lysU-luxCDABE</i>	1.3	1.4	6.0	1.3	1.3	0.89